FIG. IA

$$c = (1,1,1,1,1,0,0,0,0,0,0)$$

$$gene_1 = (e_1, e_2, e_3, \quad . \quad . \quad . \quad , e_{12})$$
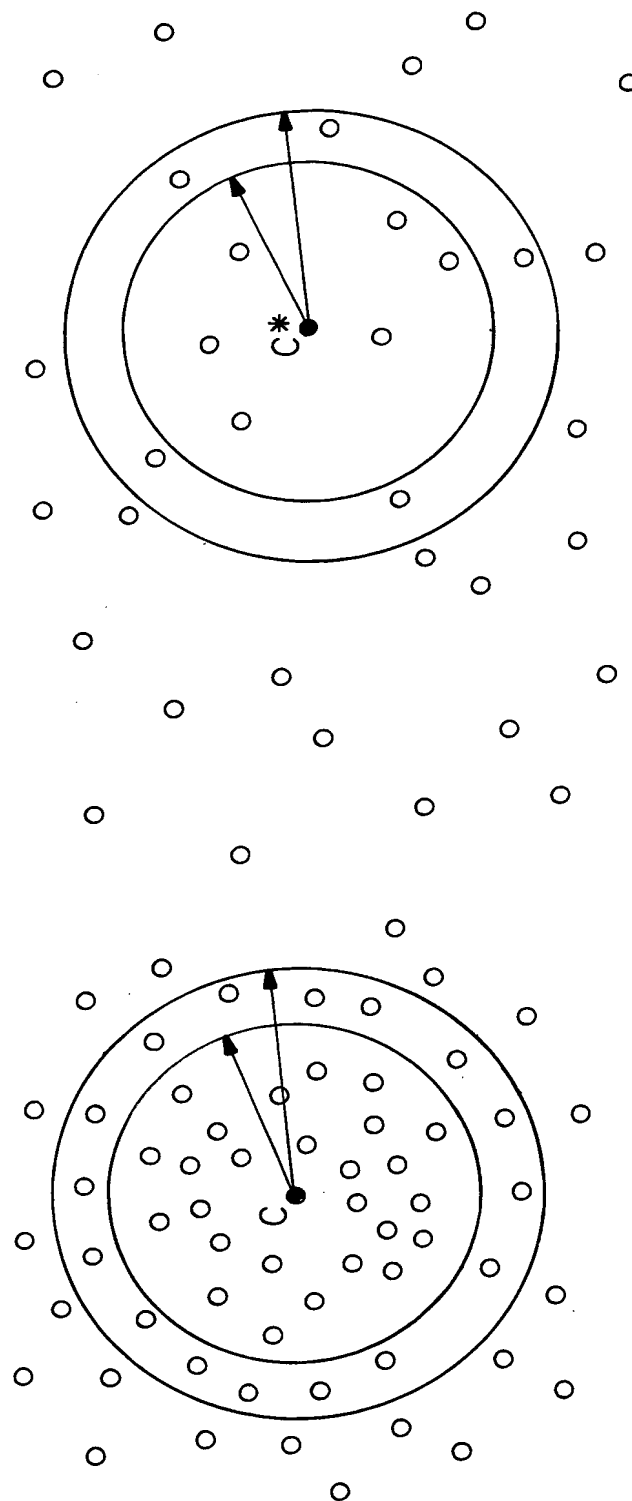
$$gene_1 = (e_1, e_2, e_3, \quad . \quad . \quad . \quad , e_{12})$$

## FIG. IB
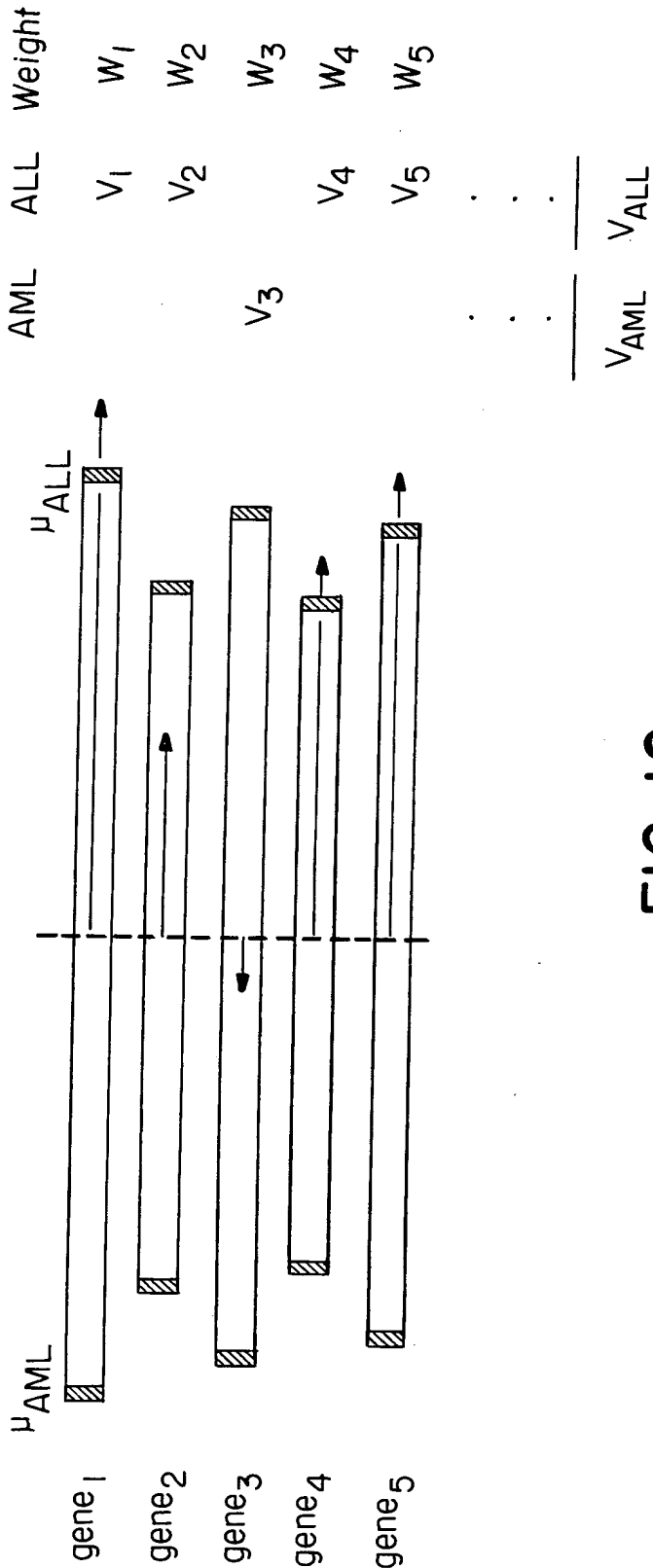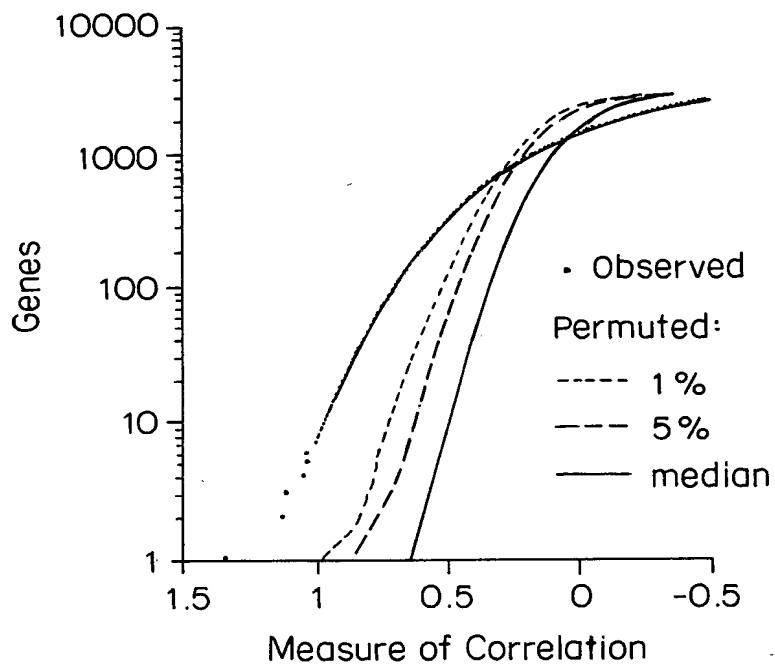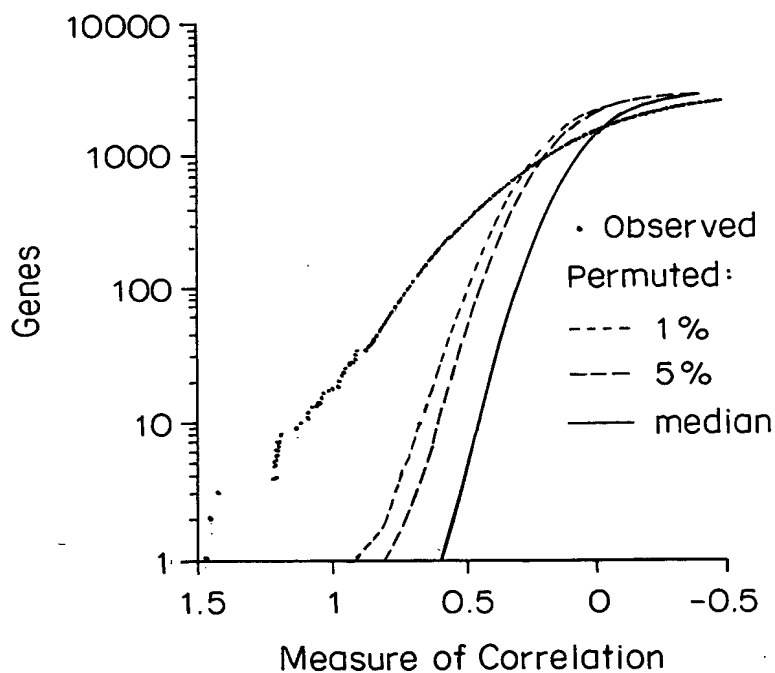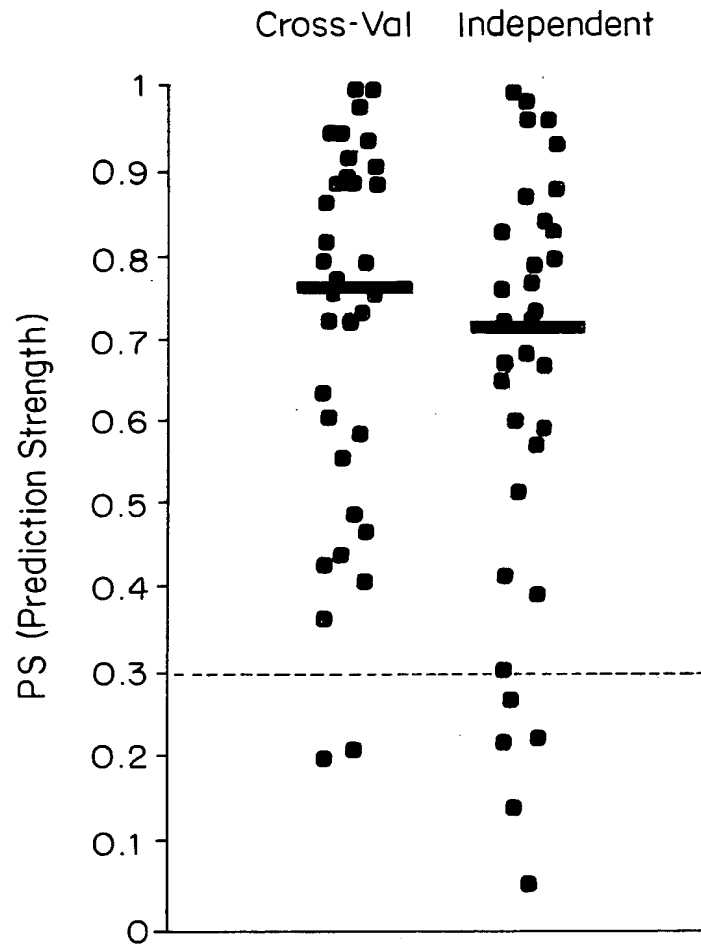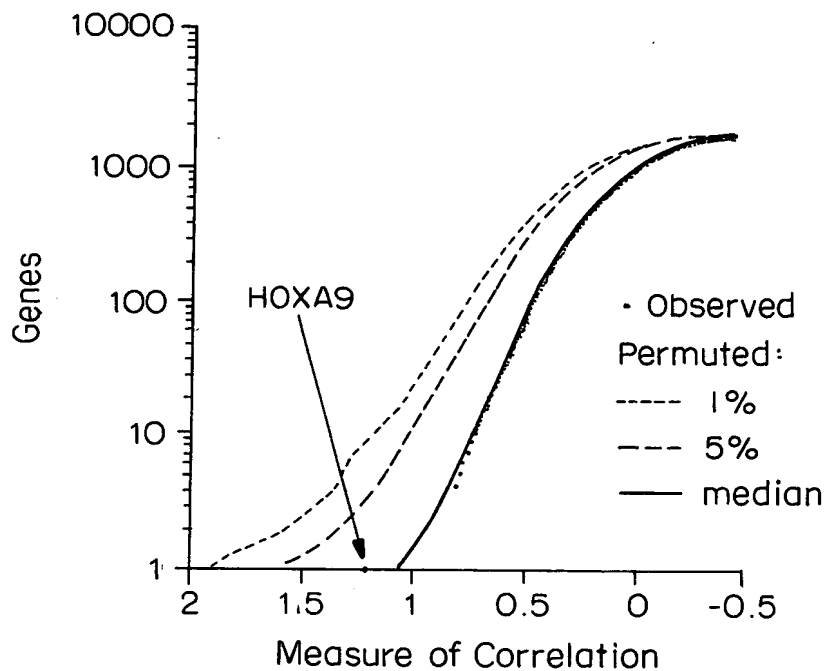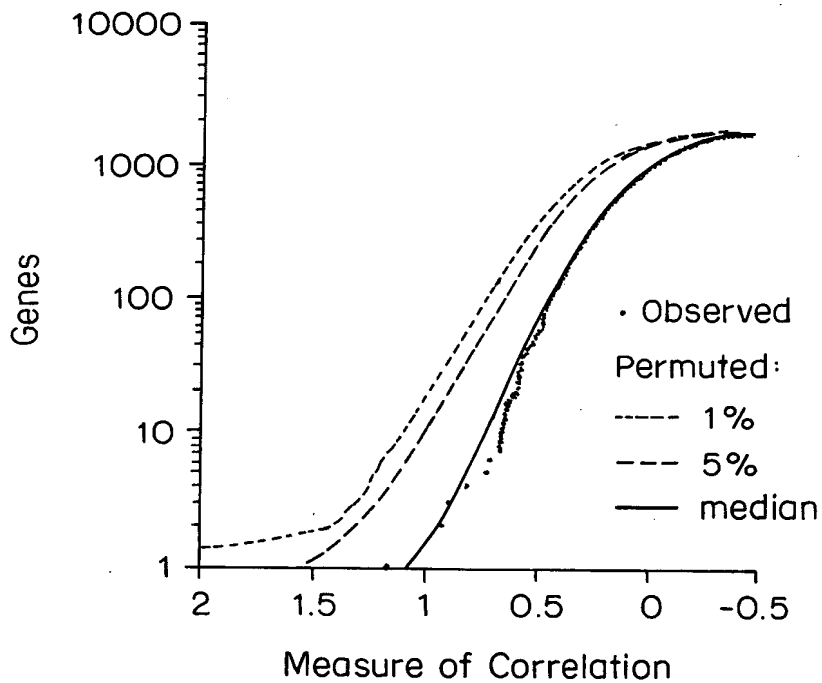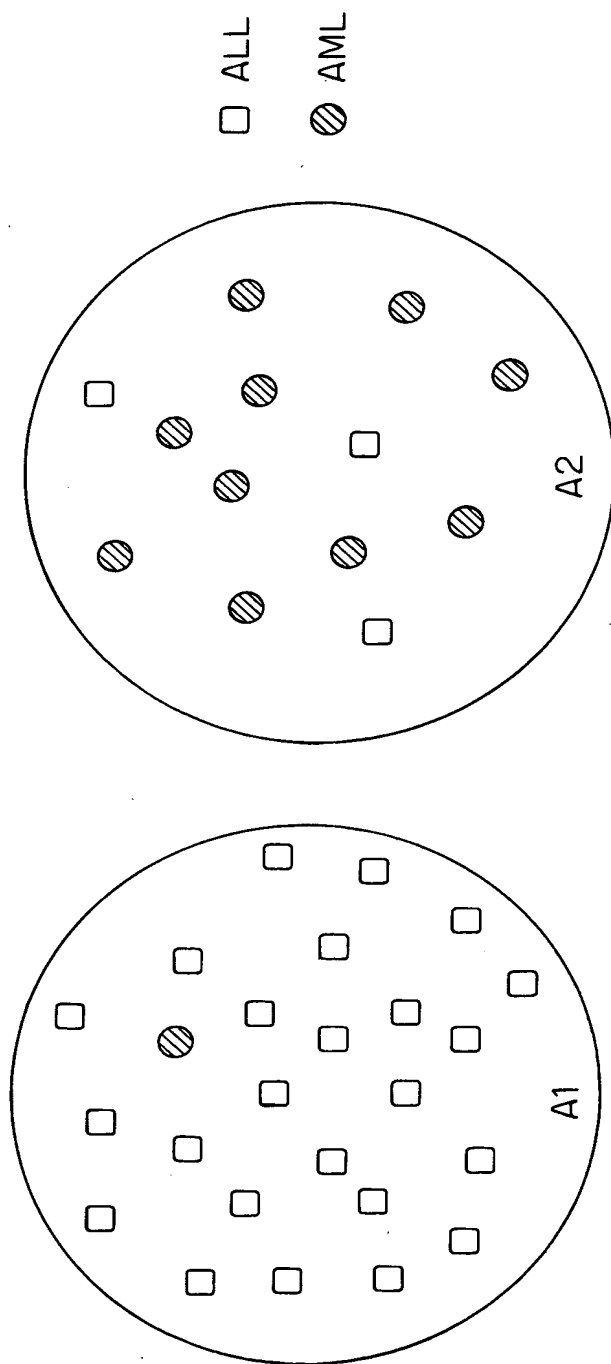
FIG. IC

FIG. 2A



FIG. 2B

Cross-Val    Independent



FIG. 3A

FIG. 4A



FIG. 4B

ALL ☐ ☐

AML ◎ ◎

A2

A1

# FIG. 5A

FIG. 5B

App No.:        10/074,789
le:    "Methods for Classifying Samples ...
ventors:       Todd R. Golub, *et al.*



◎ AML    ▨ T-ALL    ☐ B-ALL

## FIG. 5C

FIG. 5D

App No.: 10/074,789
tle: "Methods for Classifying Samples .
Inventors: Todd R. Golub, *et al.*

Input Source 17

I/O Process
19

11

13

15

Storage
21

I/O Process
23

Output

FIG. 6

Discovered Classes:

⊘,◉ Large B-Cell    ○ Follicular

# FIG. 7

Discovered Classes:



⊘ Medulloblastoma   ○ Glioma
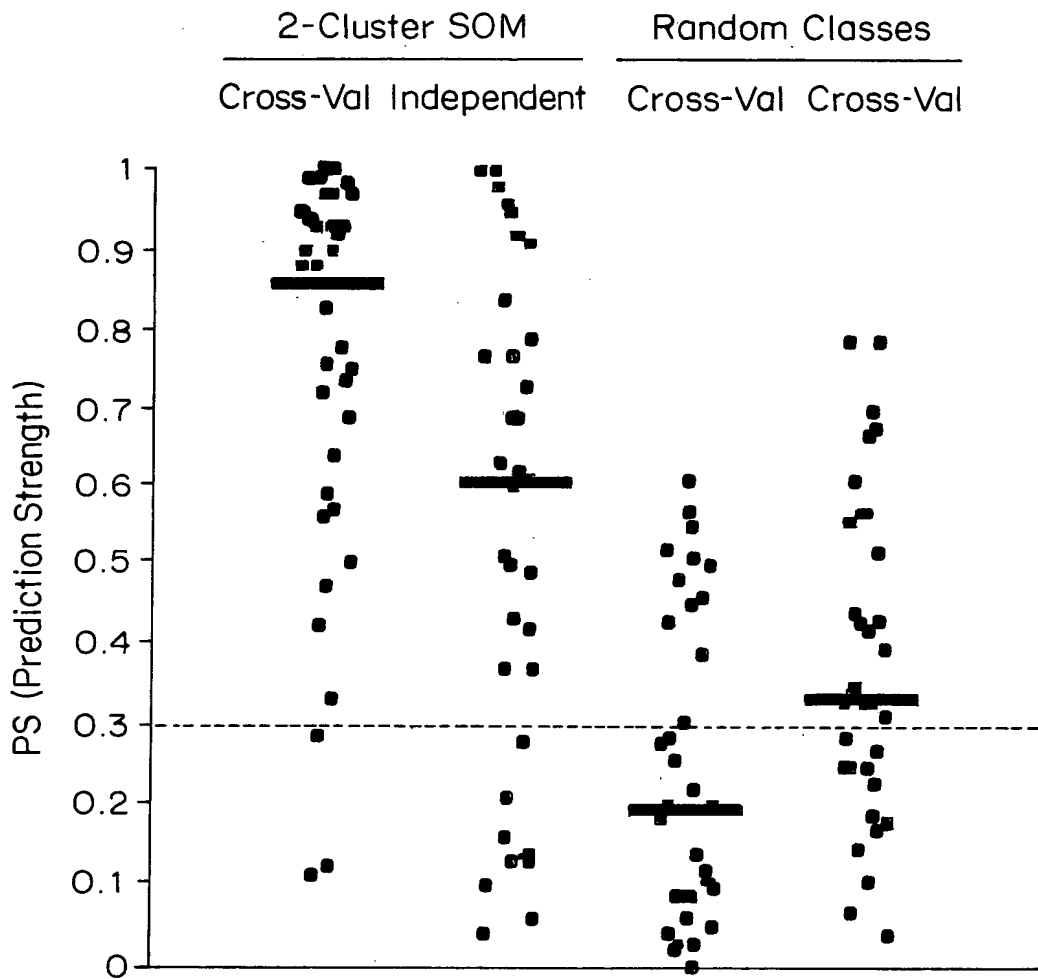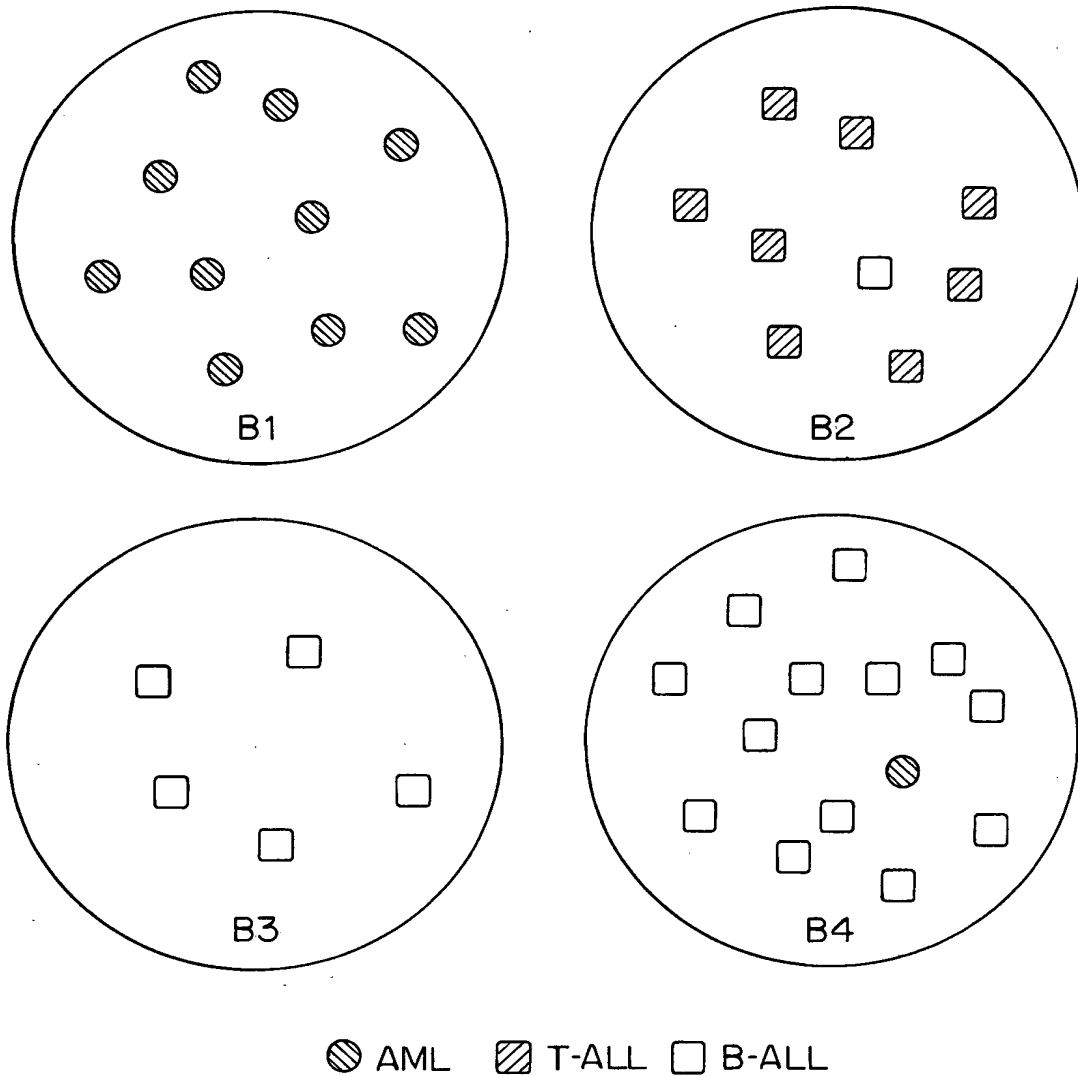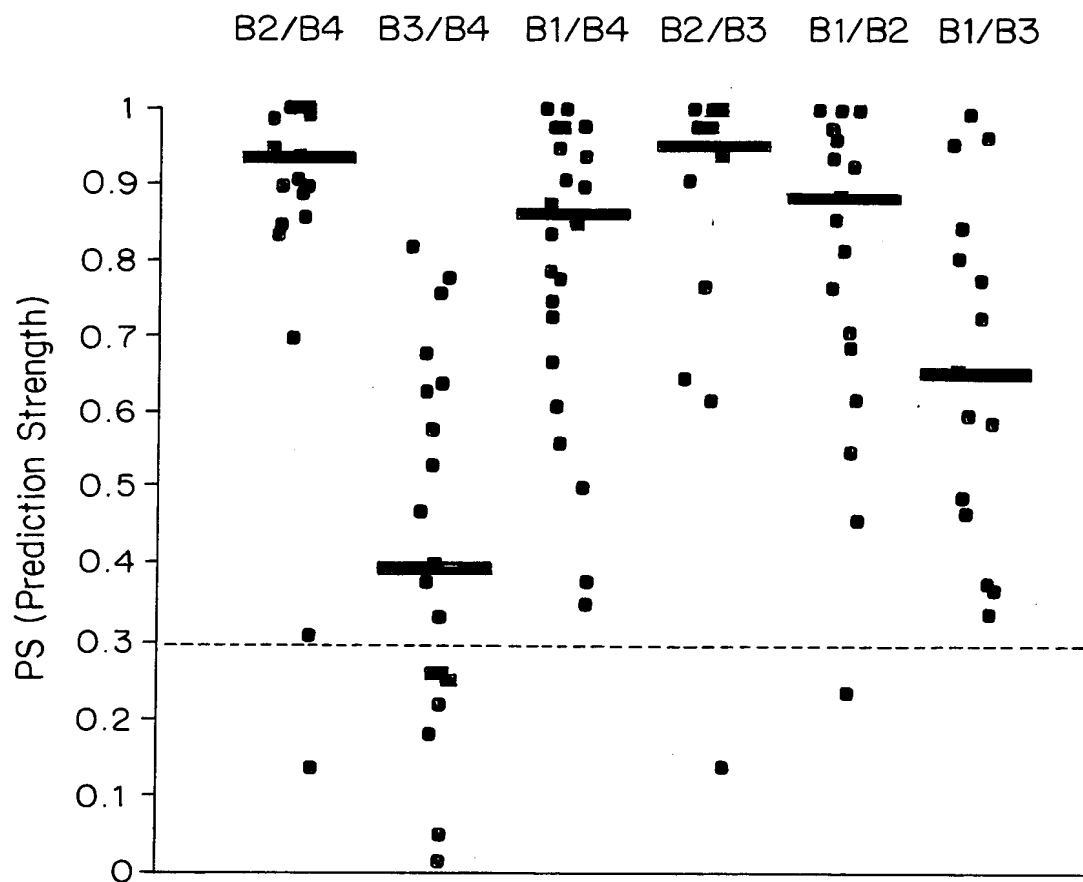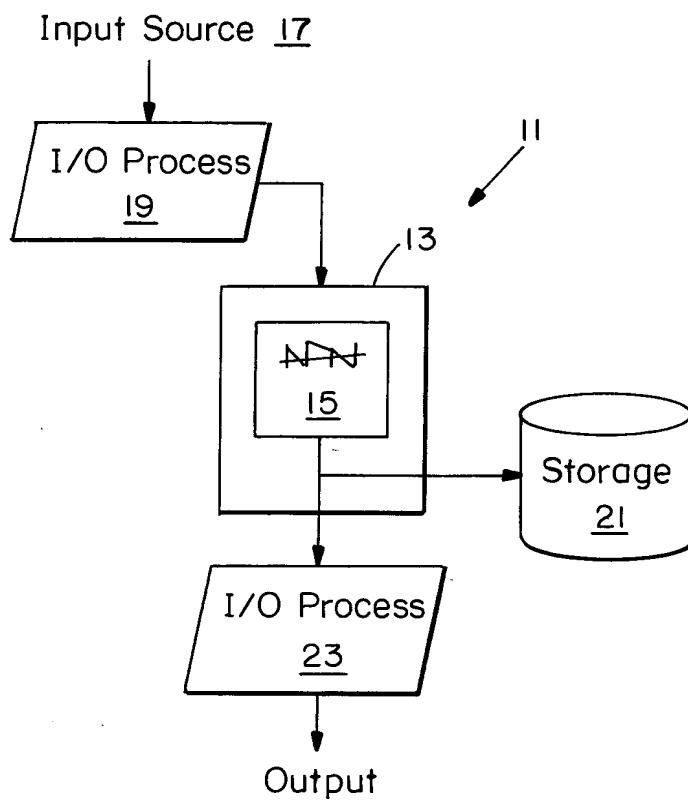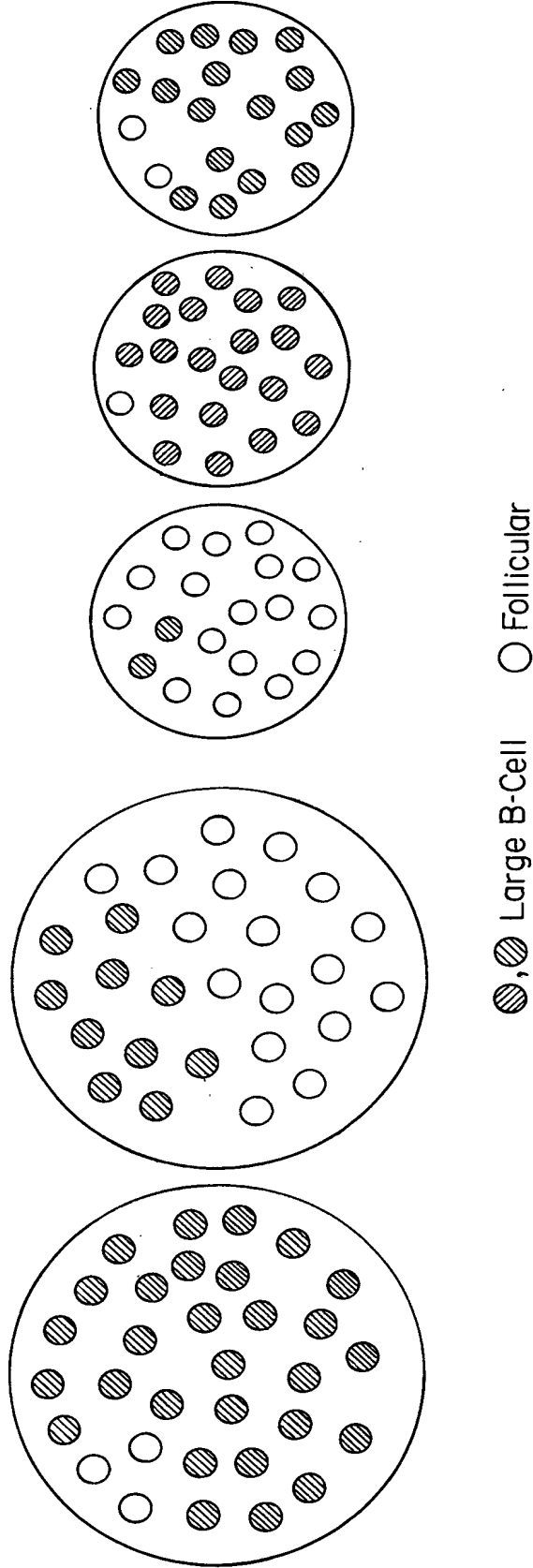
# FIG. 8

FIG. 9
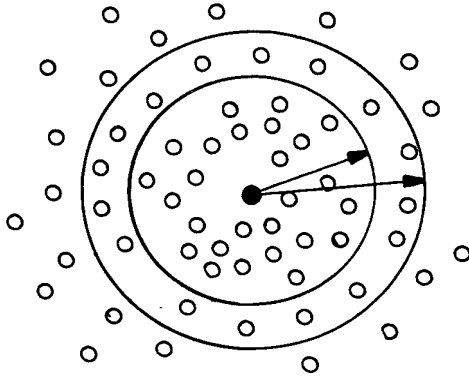
Class Pattern
Neighborhood

# FIG. 11A
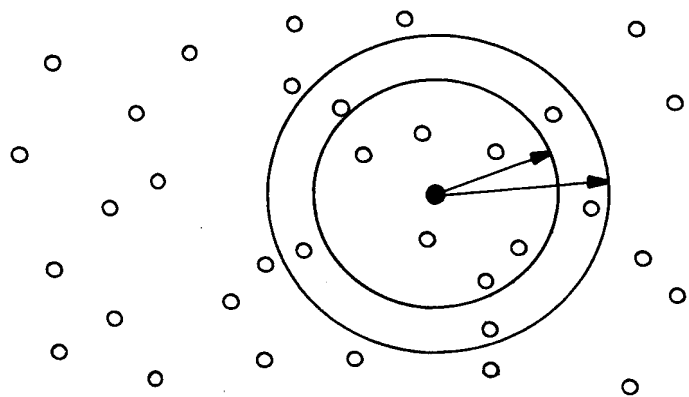
Permuted Pattern
Neighborhood

# FIG. 11B

$k = 1$

Freq.

Density around
Permuted Pattern

Ref.  5%   median    95%

Measure of Correlation

# FIG. 11C

Ref.  5%  median  95%

$k$

Significant
Neighbors

Measure of Correlation
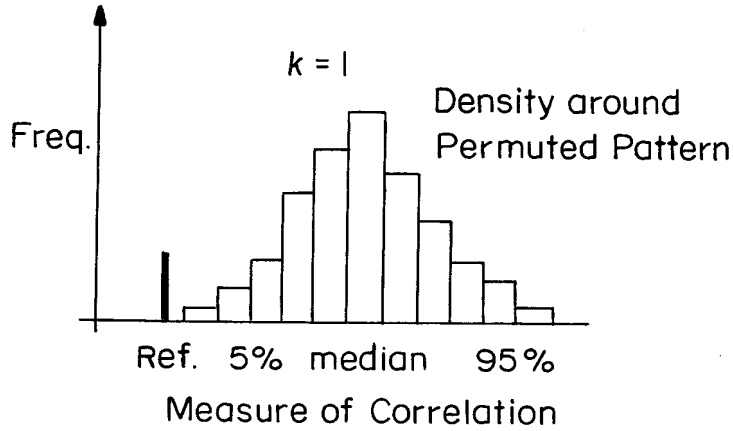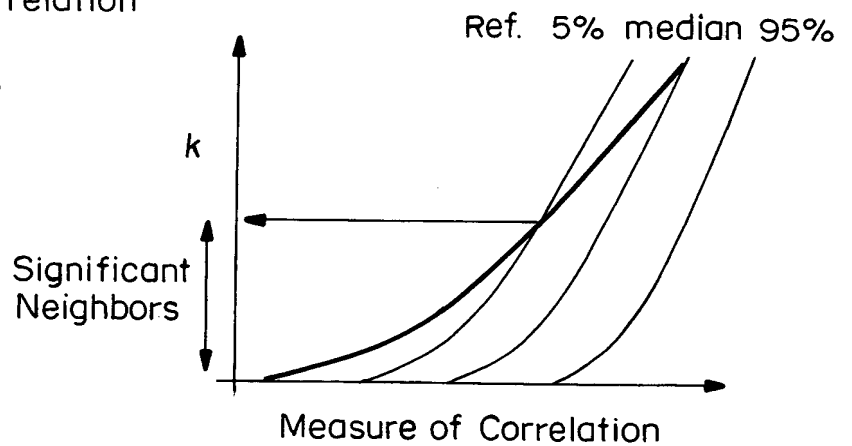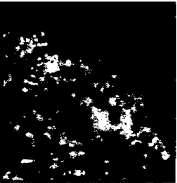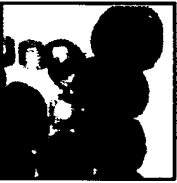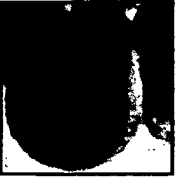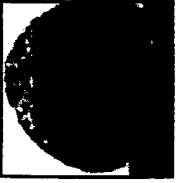
# FIG. 11D

# FIG. 10

## Hierarchy of Problems in Molecular Class Prediction

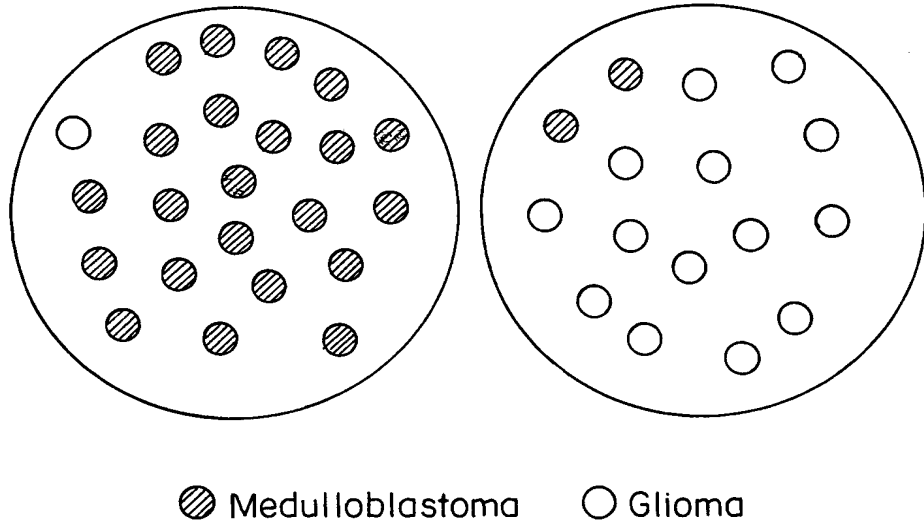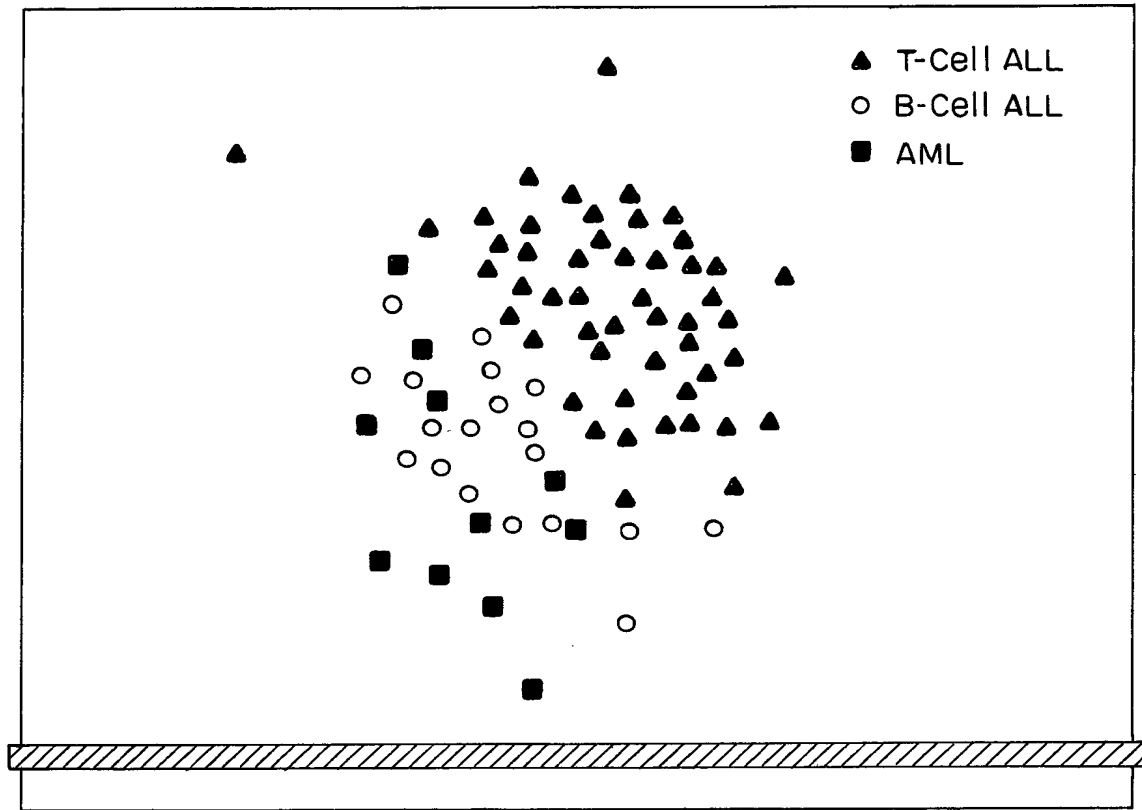| Problem: | Difficulty: | Gene Markers: | Error: | Example: |
|---|---|---|---|---|
| I. Tissue or Cell Type Normal vs. Abnormal | Low | ~1000-2000 | ~0% | Normal vs. Renal Carcinoma |
| II. Morphological Type | Low-medium | ~200-500 | ~0-5% | Leukemia ALL vs. AML |
| III. Morphological Subtype | Medium-high | ~50-100 | ~0-15% | ALL B- vs. T-Cell |
| IV. Treatment Outcome Drug Sensitivity | High | ~1-20 | ~5-50% | AML Treatment Outcome |

ALL          AML



C-myb (U22376)
Proteasome iota (X59417)
MB-1 (U05259)
Cyclin D3 (M92287)
Myosin light chain (M31211)
RbAp-48 (X74262)
SNF2 (D26156)
HkrT-1 (S50223)
E2A (M31523)
Inducible protein (L47738)
Dynein light chain (U32944)
Topoisomerase II β (Z15115)
IRF2 (X15949)
TFIIEβ (X63469)
Acyl-Coenzyme A dehydrogenase (M91432)
SNF2 (U29175)
(Ca2+)-ATPase (Z69881)
SRP9 (U20998)
MCM3 (D38073)
Deoxyhypusine synthase (U26266)
Op 18 (M31303)
Rabaptin-5 (Y08612)
Heterochromatin protein p25 (U35451)
IL-7 receptor (M29696)
Adenosine deaminase (M13792)

Fumarylacetoacetate (M55150)
Zyxin (X95735)
LTC4 synthase (U50136)
LYN (M16038)
HoxA9 (U82759)
CD33 (M23197)
Adipsin (M84526)
Leptin receptor (Y12670)
Cystatin C (M27891)
Proteoglycan 1 (X17042)
IL-8 precursor (Y00787)
Azurocidin (M96326)
p62 (U46751)
CyP3 (M80254)
MCL1 (L08246)
ATPase (M62762)
IL-8 (M28130)
Cathepsin D (M63138)
Lectin (M57710)
MAD-3 (M69043)
CD11c (M81695)
Ebp72 (X85116)
Lysozyme (M19045)
Properdin (M83652)
Catalase (X04085)

-3  -2.5  -2  -1.5  -1  -0.5  0  0.5  1  1.5  2  2.5  3
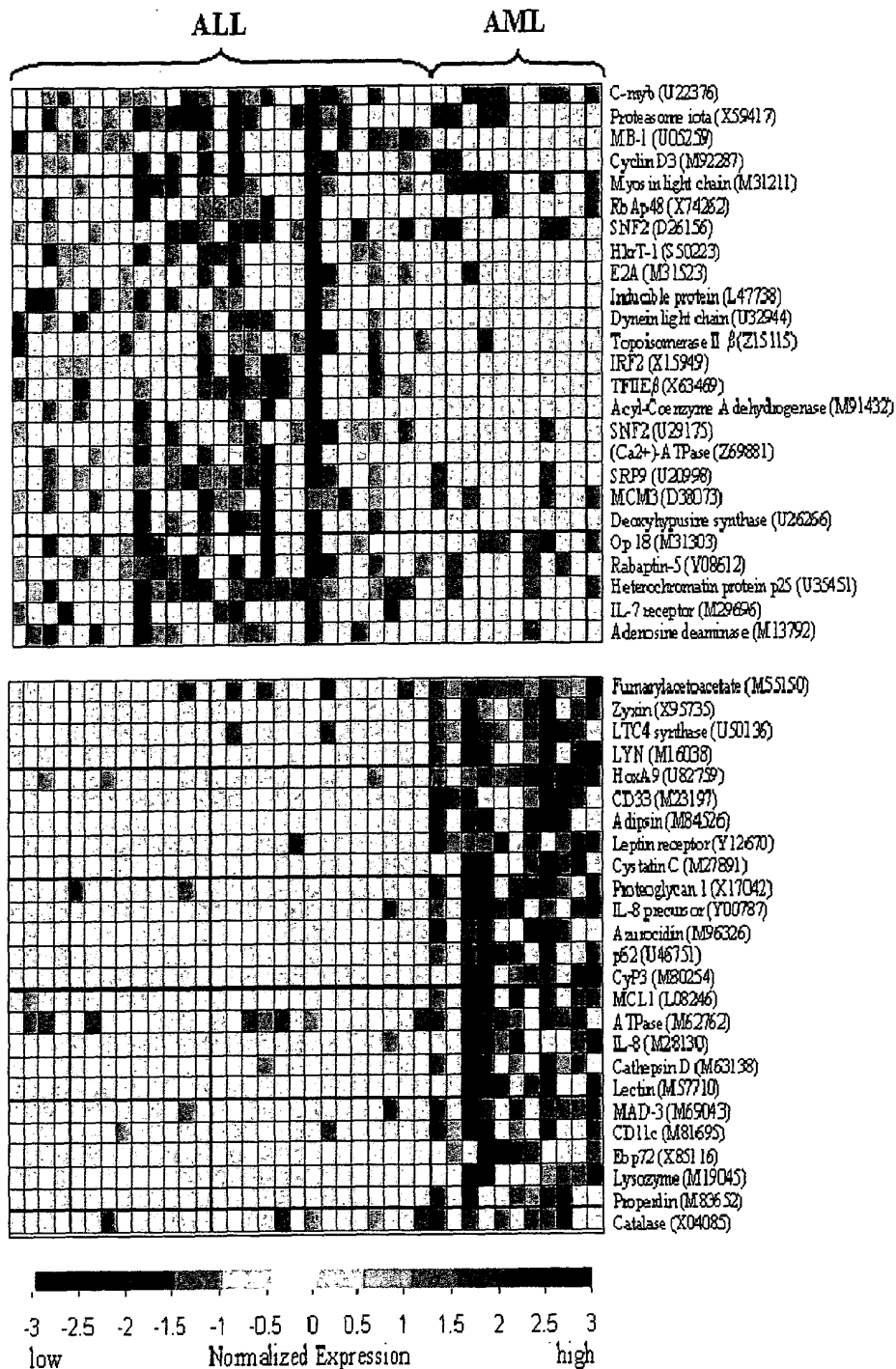low          Normalized Expression          high

FIG. 3B

CLASS PREDICTION RESULTS

| Problem Type | Biological System | Problem Distinction | Number of samples | Number of errors | Number of no calls | Number of errors (all calls) | Number of gene markers |
|---|---|---|---|---|---|---|---|
| I | Renal | Normal v. Carcinoma | 12 | 0 (0%) | 0 | 0 (0%) | >100 |
| II | Leukemia | ALL vs AML | 35 | 0 (0%) | 2 | 0 (0%) | 700 |
| III | Leukemia | ALL B v. T-Cell | 33 | 0 (0%) | 1 | 1 (3%) | 200 |
| IV | Leukemia | Treatment Outcome | 15 | 2 (13%) | 0 | 2 (13%) | 1 |

Fig. 12